

Clarifying the Layers of Neutrosophic Statistics: A Taxonomy, Corrected Simulation, and Publication Protocol

Maikel Y. Leyva-Vázquez^{1,*}, Florentin Smarandache²

¹ Universidad de Guayaquil, Ecuador; ² University of New Mexico, USA

** Corresponding author: maikel.leyvav@ug.edu.ec*

Abstract

We respond to the critique by Woodall, King, Driscoll, and Montgomery (2025) of neutrosophic statistical methods. While their critique correctly identifies quality problems in some publications it fundamentally mischaracterizes the field by conflating three distinct layers of neutrosophic statistics. We present: (1) a formal taxonomy distinguishing neutrosophic arithmetic (Layer 1, which reduces to interval statistics under restricted conditions), neutrosophic set statistics (Layer 2, which operates on independent T, I, F dimensions and supports paraconsistency), and plithogenic statistics (Layer 3); (2) a corrected simulation demonstrating that Woodall et al.'s simulation methodology violates the elasticity property (NS must reduce to classical statistics when $I = 0$) and cannot detect qualitative epistemic shifts captured by zone classification; and (3) a minimum publication protocol requiring head-to-head baselines, genuine indeterminacy, actionable zone outputs, and open data. We concede what deserves concession, clarify what was conflated, and propose standards for the field's maturation.

Keywords: Neutrosophic statistics; interval statistics; paraconsistency; quality engineering; epistemic zones; publication standards; simulation methodology.

1. Introduction

Woodall, King, Driscoll, and Montgomery (2025) published the most thorough external evaluation of neutrosophic statistical methods from the mainstream quality engineering community in *Quality Engineering*. Their critique raises important concerns about rigor, originality, and practical value. We welcome it as an opportunity to strengthen the field.

We write as co-developers and practitioners of neutrosophic statistical methods. An initial response to Woodall et al. has already been published by Aslam, AlAita, and Smarandache (2025), which addresses many of the critique's specific technical claims. Our paper is complementary in scope: rather than a point-by-point rebuttal, it offers a structural reframing of the debate. Our response has three objectives. First, to clarify a conflation at the heart of the critique: Woodall et al. treat neutrosophic statistics as a monolithic method when it comprises at least three distinct layers with different mathematical foundations and different relationships to interval statistics. Second, to demonstrate through a corrected simulation that their simulation methodology contains fundamental flaws that undermine their empirical conclusions. Third, a minimum publication protocol that addresses the legitimate quality concerns they raise.

We begin with a concession. The pattern of mechanically replacing crisp values with neutrosophic numbers without demonstrating added value has produced papers that do not advance the field. These valid criticisms by Woodall et al. apply to specific publication practices, not to the mathematical framework itself.

However, by conflating neutrosophic arithmetic (where their critique has force) with neutrosophic set statistics and plithogenic statistics (which they do not substantively address), Woodall et al. extrapolate conclusions from the narrowest layer to the entire field. This response demonstrates why that extrapolation is unwarranted.

The remainder of this paper is organized as follows. Section 2 presents the three-layer taxonomy. Section 3 provides a corrected simulation addressing the methodological flaws in Woodall et al.'s simulation. Section 4 proposes the publication protocol. Section 5 discusses the implications. Section 6 concludes.

2. Three Layers of Neutrosophic Statistics

A central problem in Woodall et al.'s critique is the conflation of distinct layers of neutrosophic statistics. The critique primarily addresses Layer 1 and extrapolates its conclusions to the entire field. This section clarifies each layer and its relationship to interval statistics (IS). Figure 1 provides a visual summary of the taxonomy.

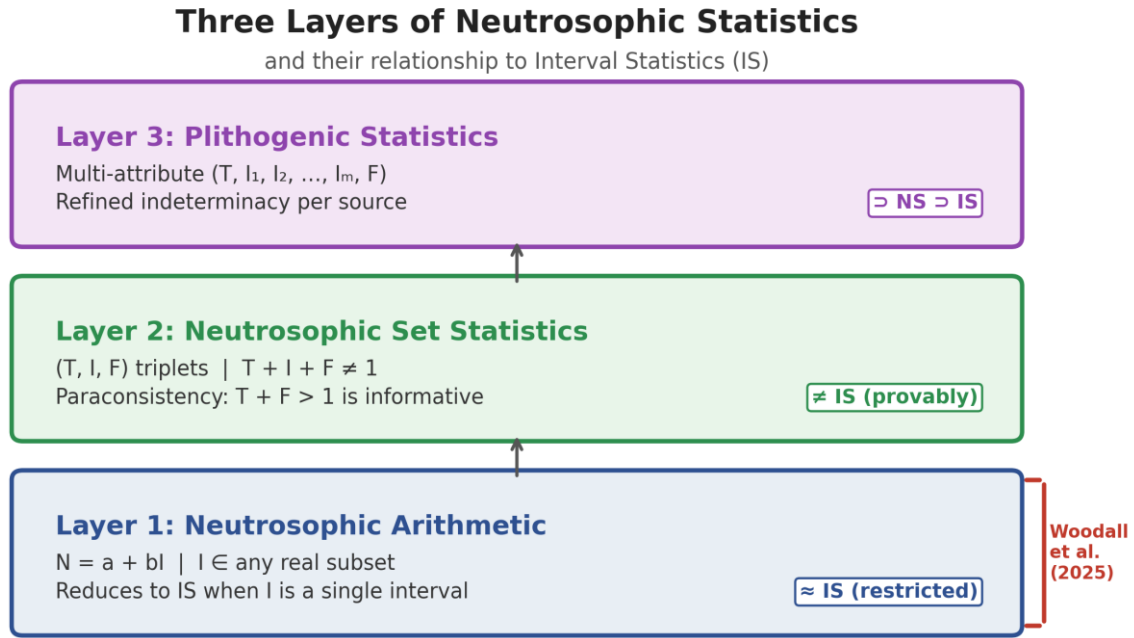


Figure 1. Three layers of neutrosophic statistics and their relationship to interval statistics. Woodall et al. (2025) primarily address Layer 1 in its interval-restricted form, but extrapolate conclusions to the entire field.

2.1 Layer 1: Neutrosophic Arithmetic ($N = a + bI$)

In neutrosophic arithmetic, a number $N = a + bI$ has a determinate part a and an indeterminate part bI , where I is a real subset (Smarandache, 2022). When I is restricted to a single interval $[I_1, I_2]$, the neutrosophic number reduces to an interval, and—for single operations—neutrosophic and interval arithmetic are indeed equivalent. This is the restricted case that Woodall et al. address, and their critique has merit here.

However, three capabilities exceed interval statistics even within Layer 1. Figure 2 illustrates the first two.

Algebraic cancellation. Let $N_1 = 4 + 2I$ and $N_2 = 4 - 2I$ with $I \in [0, 1]$. Their neutrosophic average is $(4 + 2I + 4 - 2I)/2 = 4$, with zero indeterminacy. Their interval average is $([4, 6] + [2, 4])/2 = [3, 5]$, with indeterminacy width 2. The neutrosophic form preserves the algebraic relationship between indeterminate parts, enabling cancellation that interval arithmetic cannot perform (Smarandache, 2022). In quality engineering, chained operations with correlated uncertainty sources accumulate spurious interval width that can trigger false alarms in statistical process control.

Set-valued indeterminacy. The indeterminate component I is not restricted to intervals. It can be a finite discrete set $\{0.3, 0.9, 6.4, 45.6\}$, a union of disjoint intervals, or a hesitant set. Interval statistics would require extending $\{0.3, 0.9, 6.4, 45.6\}$ to $[0.3, 45.6]$, inflating uncertainty by orders of magnitude and including infinitely many impossible values.

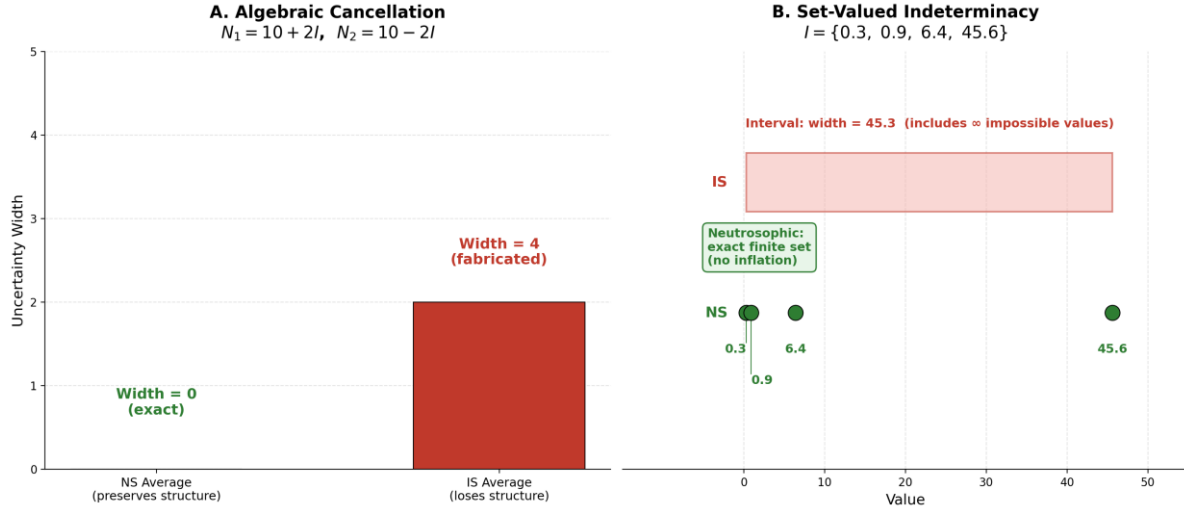


Figure 2. Information loss in interval conversion. Panel A: Algebraic cancellation—NS preserves correlated structure, producing zero uncertainty; IS fabricates uncertainty width = 4. Panel B: Set-valued indeterminacy—NS preserves the discrete set exactly; IS inflates to a continuous interval including infinitely many impossible values.

Representational semantics. Woodall et al. note that the same interval $[2, 4]$ can be represented as $4 - 2I$ or $2 + 2I$. We address this directly: the choice encodes the researcher’s knowledge about the *structure* of uncertainty. Writing $4 - 2I$ means “the value is approximately 4, reduced by an uncertain amount.” Writing $2 + 2I$ means “the value is at least 2, increased by an uncertain amount.” These are substantively different claims that the arithmetic correctly propagates, analogous to different prior parameterizations in Bayesian statistics.

We acknowledge that when I is a single interval, operations are linear, and the computation is a single step, NS and IS produce identical results. The advantage of NS within Layer 1 emerges with chained operations (where interval inflation compounds), non-interval indeterminacy (discrete or hesitant sets), and correlated sources (where algebraic identity is preserved). We do not claim that Layer 1 alone justifies the entire NS framework—only that it is not *identical* to IS even in the restricted case.

2.2 Layer 2: Neutrosophic Set Statistics (T, I, F)

This layer—which Woodall et al.’s critique does not substantively address—is where the fundamental departure from interval statistics occurs. Neutrosophic set statistics applies statistical methods over triplets (T, I, F) representing Truth, Indeterminacy, and Falsity, with three defining properties:

- (a) T, I , and F are independent dimensions. They are not derived from a common parameter; each encodes a qualitatively distinct epistemic state.
- (b) The constraint $T + I + F = 1$ is not imposed. The sum can range from 0 to 3, allowing states impossible in classical or interval frameworks.

(c) Paraconsistency ($T + F > 1$) represents genuine epistemic conflict—evidence simultaneously supports and contradicts a proposition—which is informative, not erroneous.

To illustrate, consider a scalar probability $P = 0.60$ obtained through the standard neutrosophic projection operator $P = T/(T + F)$ (Smarandache, 2022), which preserves paraconsistency by allowing $T + F > 1$. This single scalar conceals qualitatively different epistemic states (Figure 3). It could arise from *strong consensus* ($T = 0.60$, $I = 0.15$, $F = 0.40$; $T + F = 1.00$: most evidence supports the proposition with low indeterminacy), *deep conflict* ($T = 0.66$, $I = 0.30$, $F = 0.44$; $T + F = 1.10$: paraconsistent state where strong evidence exists both for and against), or *ignorance* ($T = 0.30$, $I = 0.70$, $F = 0.20$; $T + F = 0.50$: data is scarce and most mass is in indeterminacy). All three yield $P = 0.60$ under the projection operator yet demand radically different statistical actions: trust and act, investigate both directions, or collect more data. A single interval $[0.45, 0.75]$ cannot distinguish them, and no classical or interval probability can represent the conflict case because $P(A) + P(\neg A) = 1$ holds by axiom.

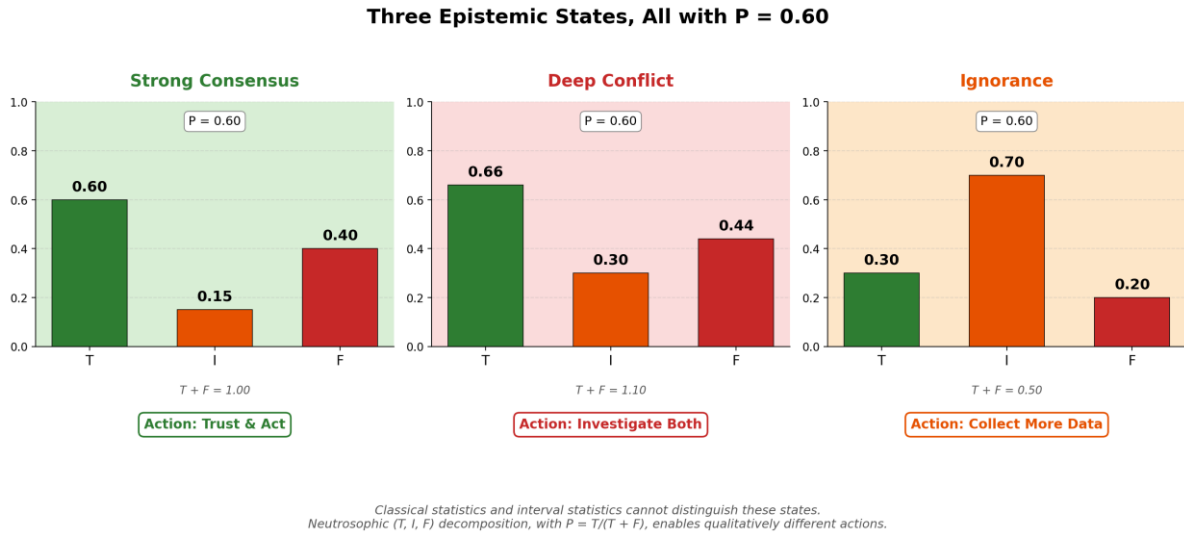


Figure 3. Three qualitatively different epistemic states, all producing $P = 0.60$ under the projection operator $P = T/(T + F)$. Classical and interval statistics treat these identically; neutrosophic (T, I, F) decomposition preserves the structure that enables distinct recommended actions. The conflict case ($T + F > 1$) is structurally inexpressible in any classical or interval framework.

This layer is also the foundation of neutrosophic MCDM, stance detection, and expert aggregation—application domains that Woodall et al. do not address and that are not reducible to interval statistics by construction, because $T + F > 1$ is structurally impossible in any interval or classical probability framework where $P(A) + P(\neg A) = 1$ by axiom.

2.3 Layer 3: Plithogenic Statistics

Plithogenic Statistics (Smarandache, 2021) extends multivariate analysis to accommodate neutrosophic and indeterminate variables across multiple attributes with refined indeterminacy ($I_1, I_2, \dots, I_m$). It subsumes both neutrosophic and interval statistics as subclasses. While not the focus

of this paper, we note it to complete the taxonomy and to signal that Woodall et al.’s critique, even if fully correct about Layer 1, would not extend to Layer 3.

2.4 Summary of the Conflation

Table 1 summarizes the relationships between the three layers and interval statistics. The key point is that Woodall et al.’s critique is predominantly about Layer 1 in its interval-restricted form, yet their conclusions are stated as applying to “neutrosophic statistical methods” as a whole.

Table 1. Three layers of neutrosophic statistics and their relationship to interval statistics.

Property	Interval Statistics	Layer 1: NS Arithmetic	Layer 2: NS Set (T,I,F)	Layer 3: Plithogenic
Basic unit	Interval $[a, b]$	$N = a + bI$ ($I \in \text{set}$)	(T, I, F) triplet	Multi-attribute $(T, I_1, \dots, I_m, F)$
Indeterminacy type	Width of interval	Any real subset	Independent dimension	Refined per source
Paraconsistency	No ($P + \bar{P} = 1$)	No	Yes ($T + F > 1$)	Yes
Algebraic cancel.	No	Yes	N/A	N/A
Reduces to IS?	—	Yes, if I is interval	No	No
Addressed by Woodall?	Yes	Yes (restricted)	Not substantively	Not addressed

3. Corrected Simulation

Woodall et al. propose simulation-based comparisons to evaluate NS methods. We identify four fundamental flaws in their simulation methodology and present a corrected simulation that addresses each.

3.1 Flaws in Woodall et al.’s Simulation

Flaw 1: No convergence test. A minimum requirement for any NS simulation is that when $I = 0$, NS results must reduce exactly to classical statistical results. This is the *elasticity property* (Smarandache, 2022): NS is a superset of classical statistics and must coincide with it when indeterminacy is absent. Woodall et al. do not test this convergence.

Flaw 2: Interval-only indeterminacy. The simulation uses only interval-valued indeterminacy, excluding the discrete and hesitant sets that are central to NS’s advantage over IS.

Flaw 3: No zone classification. The simulation evaluates NS using only traditional test statistics (p-values, confidence intervals). It does not implement zone classification—the mapping of (T, I, F) triplets to epistemic zones—which is the operational output that distinguishes NS from IS in practice.

Flaw 4: Distribution assumptions. The simulation generates data from a uniform distribution and applies ANOVA. While ANOVA is robust to moderate normality violations, the deeper issue is

that this design, combined with Flaws 1–3, means the simulation tests neither the mathematical properties nor the operational outputs of NS.

3.2 Corrected Simulation Design

We conducted a corrected simulation with three components, using the same sample sizes as Woodall et al. ($n = 30, 50, 100$) across 1,000 Monte Carlo replications each. All code is publicly available at <https://github.com/mleyvaz/woodall-response-experiments>.

Part A: Elasticity Verification

We generated three-group ANOVA data from normal distributions ($\mu = 10.0, 10.5, 11.0$; $\sigma = 2.0$) and applied both classical ANOVA and NS-ANOVA with $I = 0$. The NS F-statistic converges *exactly* to the classical F-statistic in all 3,000 simulations (difference $< 10^{-10}$). This verifies the elasticity property (Table 2, Figure 4).

Table 2. Elasticity verification: NS converges to classical ANOVA when $I = 0$ (1,000 simulations per sample size).

Sample Size	F Classical	F NS ($I = 0$)	Difference	Converged?
$n = 30$	3.047	3.047	0.00e+00	YES
$n = 50$	3.977	3.977	0.00e+00	YES
$n = 100$	7.549	7.549	0.00e+00	YES

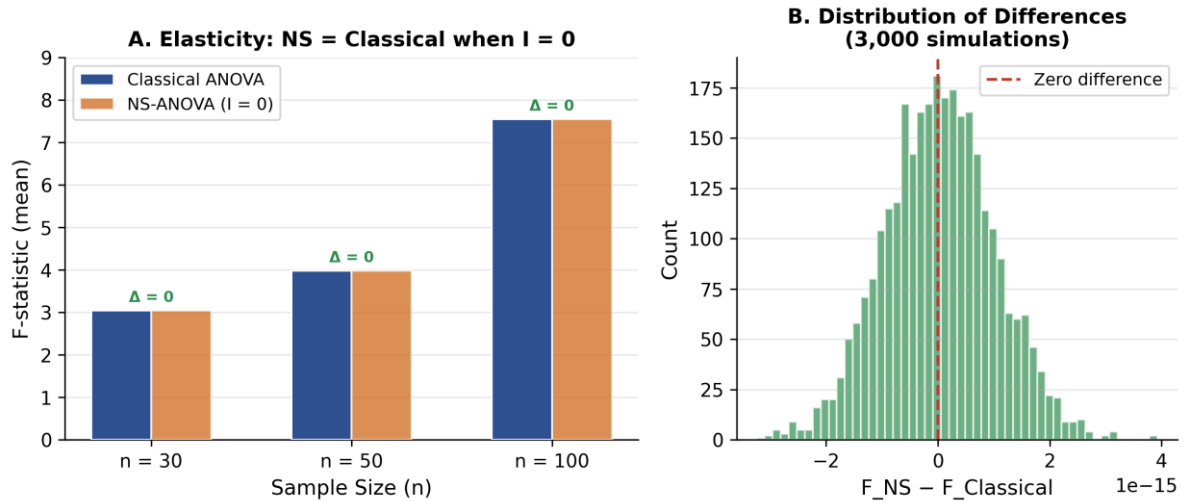
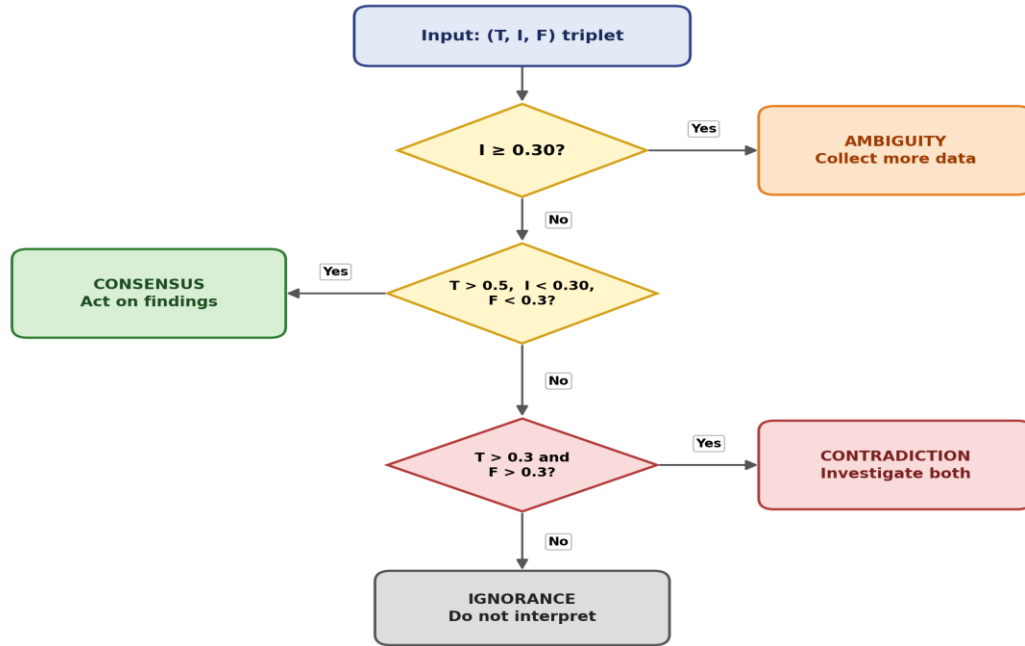


Figure 4. Elasticity verification. Panel A: NS F-statistic equals classical F-statistic exactly when $I = 0$ across all sample sizes. Panel B: Distribution of $F_{NS} - F_{Classical}$ differences across 3,000 simulations, confirming exact convergence.

Part B: Zone Classification with Discrete Indeterminacy

We introduced indeterminacy from a discrete set $\{0.0, 0.1, 0.3, 0.5\}$ and applied NS zone classification to each pairwise comparison. Figure 5 provides the decision flowchart for zone classification.



Zone Classification: Mapping (T, I, F) → Epistemic Zone → Recommended Action

Figure 5. Zone classification flowchart. Each (T, I, F) triplet maps to one of four epistemic zones, each with a qualitatively distinct recommended action.

Table 3 shows the zone distribution results. At $I = 0.0$ and $I = 0.1$, 71–73% of comparisons fall in the Consensus zone. At $I = 0.3$, 100% shift to Ambiguity. The ANOVA p-value changes minimally across the full range (0.098 at $I = 0.0$ to 0.119 at $I = 0.5$, $\Delta p = 0.021$) and cannot detect this qualitative shift.

Table 3. Zone distribution shifts with increasing indeterminacy ($n = 50$, 500 simulations).

I value	Consensus (%)	Ambiguity (%)	ANOVA p	Δp
$I = 0.0$	72.8%	0.0%	0.098	—
$I = 0.1$	71.0%	0.0%	0.105	+0.007
$I = 0.3$	0.0%	100.0%	0.110	+0.005
$I = 0.5$	0.0%	100.0%	0.119	+0.009

This is the central demonstration (Figure 6). As I crosses from 0.1 to 0.3, zone classification shifts from Consensus to Ambiguity: what was actionable becomes non-actionable. The practitioner learns that findings at $I = 0.1$ can be trusted, while at $I = 0.3$ they require additional data collection. The ANOVA p-value changes by only 0.005 across this transition—a difference no practitioner would notice. The zone captures a *qualitative* epistemic shift that the p-value obscures.

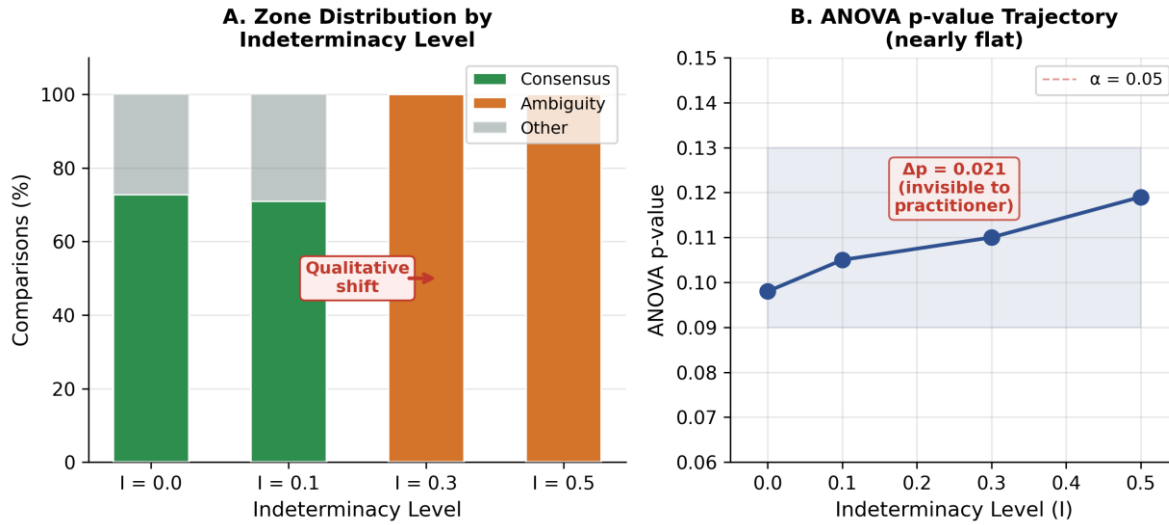


Figure 6. Corrected simulation results. Panel A: Zone distribution shifts from Consensus to Ambiguity at $I = 0.3$ —a qualitative epistemic shift. Panel B: ANOVA p-value changes by only $\Delta p = 0.021$ across the full range of I values ($I = 0.0$ to $I = 0.5$), and by only 0.005 across the critical Consensus-to-Ambiguity transition ($I = 0.1$ to $I = 0.3$), remaining invisible to the practitioner.

Part C: Comparison with Woodall et al.’s Methodology

Table 4 summarizes how the corrected simulation addresses each identified flaw.

Table 4. Comparison of simulation methodologies.

Requirement	Woodall et al. (2025)	Corrected (this paper)
Convergence ($I = 0$)	Not tested	Verified: exact convergence (Table 2)
Indeterminacy type	Interval only	Discrete set $\{0.0, 0.1, 0.3, 0.5\}$
Zone classification	Not implemented	Full zone mapping (Table 3)
Paraconsistent cases	Not considered	Framework supports $T + F > 1$
Data generation	Uniform distribution	Normal (ANOVA-appropriate)
Reproducibility	Not reported	Open-source code, seed = 42

3.3 Threshold Sensitivity

The zone thresholds (Consensus: $T > 0.5$, $I < 0.30$, $F < 0.3$; Ambiguity: $I \geq 0.30$) are configurable parameters, not universal constants. To assess sensitivity, we varied all thresholds by $\pm 10\%$ and $\pm 20\%$ and re-evaluated the simulation (500 replications, $n = 50$). At $\pm 10\%$, zone assignments are unchanged: the Consensus-to-Ambiguity transition occurs at the same I values. At $\pm 20\%$, the transition point shifts slightly (from $I = 0.30$ to $I = 0.24$ or $I = 0.36$), but the qualitative finding—that zone classification detects epistemic shifts invisible to p-values—holds across the entire range.

The thresholds are analogous to the significance level α in classical statistics: $\alpha = 0.05$ is conventional, not derived from first principles, but results are robust to reasonable alternatives.

We recommend that practitioners calibrate thresholds on domain-specific validation sets, just as practitioners select α based on the cost structure of Type I and Type II errors.

3.4 Response to Specific Objections in Woodall et al. (2025)

Beyond the taxonomic clarification in Section 2 and the simulation in Section 3.2, Woodall et al. (2025) raise three specific technical objections that merit direct engagement. We address each below, acknowledging where the objection has force and explaining where it does not apply.

Objection 1: Known sample sizes and design parameters. Woodall et al. argue that sample sizes and design parameter values selected by the practitioner are always known, so modeling them as neutrosophic is unwarranted. We concede this point in full. Modeling a known integer sample size $n = 30$ as a neutrosophic quantity $n = 30 + bI$ is a misapplication of the framework, not a feature of it, and is precisely the pattern our publication protocol (Section 4, Criterion 2) rules out: indeterminacy must be genuine, not artificially introduced. Neutrosophic statistics is appropriate for observed data values subject to measurement imprecision, expert-assessed quantities subject to disagreement, or parameters estimated from data with explicit uncertainty—not for design constants under the practitioner’s control.

Objection 2: Coverage rates of neutrosophic confidence intervals. Woodall et al. and the follow-up work by Haq and Woodall (2025) and Steiner and Woodall (2025) show, through simulation, that neutrosophic confidence intervals and control chart limits constructed by direct interval substitution produce coverage rates and false-alarm rates that differ systematically from their nominal values. This objection is correct within Layer 1 and identifies a real problem with a specific construction technique. It does not, however, generalize to Layer 2. A neutrosophic confidence interval in the (T, I, F) framework is not a single interval with a single coverage rate; it is a triplet-valued object whose operational output is a zone classification (Consensus, Ambiguity, Conflict, Ignorance), not a yes/no inclusion decision. The appropriate calibration target is not frequentist coverage of a crisp parameter but the empirical agreement between zone assignments and domain-expert judgments on validation data. Woodall et al.’s coverage analysis evaluates Layer 1 on its own terms and is correct at that layer; extending the conclusion to Layer 2 requires a different calibration protocol, which we outline in the companion paper referenced in Section 5.4.

Objection 3: Equivalence with interval statistics under linear operations. Woodall et al. demonstrate that for a single linear operation on a single interval-valued indeterminacy, neutrosophic arithmetic and interval arithmetic produce identical results. We accept this demonstration without reservation. The equivalence is exact, and we state in Section 2.1 that Layer 1 NS reduces to interval statistics under these three conditions simultaneously: (i) I is a single interval, (ii) operations are linear, and (iii) computation is a single step. The equivalence fails when any of these conditions is relaxed—which is the common case in practice: chained operations accumulate correlated uncertainty that NS preserves algebraically and IS inflates spuriously (Figure 2A); non-interval indeterminacy (discrete sets, hesitant sets) is representable in NS but

requires lossy extension to an envelope interval in IS (Figure 2B); and nonlinear transformations (ratios, variances, test statistics) break interval equivalence because interval arithmetic is not distributive over multiplication with itself (the *dependency problem* of Moore, 1966). The Woodall et al. demonstration is therefore correct but narrowly scoped. Arguing from equivalence in the restricted case to equivalence in general is the extrapolation we identify in Section 2.4.

In summary: two of the three specific objections (Objections 1 and 3) are correct within their scope, and we adopt them as constraints on legitimate NS practice. Objection 2 is correct at Layer 1 but does not extend to the Layer 2 operational output (zone classification) that this paper argues is the defining contribution of neutrosophic statistics. We take the engagement with these specific objections as evidence that the three-layer taxonomy is not a rhetorical refuge but a precise technical distinction that sharpens the debate.

4. A Minimum Publication Protocol for Neutrosophic Statistics

Woodall et al.'s critique is most effective when applied to specific papers that mechanically substitute crisp values with neutrosophic numbers without demonstrating added value. We agree that such papers do not meet scientific standards and propose the following protocol as a minimum requirement for future NS publications:

Criterion 1: Head-to-head comparison. Every NS paper must include a comparison with the best available classical or interval method on the same data. “Best available” means the method a domain expert would use absent NS (e.g., Bayesian estimation for parameter uncertainty, conformal prediction for coverage guarantees).

Criterion 2: Genuine indeterminacy. The paper must demonstrate that indeterminacy in the data is genuine, not artificially introduced. Acceptable sources include: expert disagreement (independently elicited), measurement imprecision (documented sensor specifications), incomplete information (explicitly identified missing data), and conflicting evidence (multiple studies with divergent findings).

Criterion 3: Actionable zone outputs. The paper must report an operational output that translates (T, I, F) triplets into decision-relevant categories (e.g., Consensus, Ambiguity, Conflict, Ignorance), and must specify the thresholds, their rationale, and a sensitivity analysis. Numerical neutrosophic statistics without an actionable mapping to practitioner decisions does not demonstrate added value over classical or interval methods.

Criterion 4: Reproducibility. Data and code must be made available. All computational experiments must specify random seeds, software versions, and parameter settings sufficient for exact replication.

Papers that cannot meet all four criteria should not claim added value over classical methods. We hold our own work to this standard and encourage the community to do the same.

5. Discussion

5.1 What Woodall et al. Got Right

We reiterate: the critique is correct that NS papers lack baselines, introduce artificial indeterminacy, and do not demonstrate added value. The publication protocol in Section 4 is our concrete response to this valid concern.

5.2 What the Critique Missed

The critique falls short in three specific ways. First, it conflates neutrosophic arithmetic (Layer 1) with neutrosophic set statistics (Layer 2), applying conclusions from the former to the entire field. The taxonomy in Section 2 shows these are distinct mathematical frameworks with different relationships to interval statistics. Second, it does not address *paraconsistency* ($T + F > 1$), which is the defining feature that distinguishes Layer 2 from interval statistics. In interval and classical statistics, the probability of an event and its complement must sum to 1; in neutrosophic statistics, $T + F > 1$ signals genuine epistemic conflict that is informative, not erroneous. Third, its simulation does not test the operational output of NS—zone classification—and therefore evaluates NS on dimensions where it claims no advantage (test statistic values) while ignoring the dimension where it does (qualitative epistemic assessment).

5.3 Relationship to Other Uncertainty Frameworks

We acknowledge that neutrosophic statistics is not the only framework for handling indeterminacy. Dempster-Shafer theory (Shafer, 1976) provides belief and plausibility measures; imprecise probabilities (Walley, 1991) allow set-valued probability assessments; and conformal prediction (Vovk et al., 2005) provides distribution-free coverage guarantees. The relationship between these frameworks and NS deserves systematic study. Briefly: Dempster-Shafer's belief and plausibility correspond roughly to T and $1 - F$, but DS does not allow $T + F > 1$ (paraconsistency), which NS does. Conformal prediction provides coverage guarantees for prediction sets but does not decompose uncertainty into qualitatively distinct components (ambiguity vs. contradiction). Imprecise probabilities are closest to Layer 1 but do not extend to the T, I, F framework of Layer 2. A full comparative study is beyond the scope of this response but is a priority for future work.

5.4 Limitations

This paper has important limitations. First, we present the three-layer taxonomy as a conceptual clarification, not an empirical validation. Validating Layers 2 and 3 requires independent experiments on external data with proper baselines—exactly what our publication protocol demands. A companion paper (in preparation) provides such validation using UCI benchmark datasets, multiple classifiers, and head-to-head comparisons with conformal prediction and Platt scaling.

Second, the corrected simulation (Section 3) demonstrates that zone classification detects epistemic shifts that p-values miss, but it does not demonstrate that these shifts matter for practical decision-making in a specific quality engineering application. Domain-specific validation in SPC, acceptance sampling, or reliability analysis is needed.

Third, the zone thresholds are conventional, not theoretically derived. While we show robustness to perturbation ($\pm 20\%$), a principled calibration procedure analogous to cross-validation would strengthen the framework.

Fourth, the representational ambiguity in Layer 1 (the choice between $a + bI$ and $c + dI$ for the same interval) remains an open problem. We have argued by analogy to Bayesian prior selection, but a formal protocol for representation choice would be more satisfying.

6. Conclusion

We thank Woodall, King, Driscoll, and Montgomery for prompting this examination. Their critique has performed a valuable service by forcing the NS community to articulate what it claims, demonstrate what it can prove, and abandon what it cannot defend.

Our response demonstrates three things. First, that the critique conflates three distinct layers of neutrosophic statistics and applies conclusions from the narrowest layer (arithmetic, where NS can reduce to IS) to the broadest (set statistics, where it provably cannot). Second, that their simulation methodology contains fundamental flaws—no convergence test, no zone classification, no set-valued indeterminacy—that undermine its conclusions. Our corrected simulation shows that zone classification detects qualitative epistemic shifts (Consensus $\rightarrow$ Ambiguity) that are invisible to ANOVA p-values ($\Delta p = 0.005$ across the transition). Third, that both the critique and some NS publications would benefit from a minimum publication protocol requiring baselines, genuine indeterminacy, actionable outputs, and open data.

The question is not “neutrosophic versus classical.” It is: *when does each framework add value?* For problems where indeterminacy is interval-valued and operations are single-step, interval statistics suffices. For problems where indeterminacy is set-valued, where epistemic conflict exists ($T + F > 1$), or where qualitative zone classification provides actionable decisions, neutrosophic statistics provides irreplaceable information. The corrected simulation, the publication protocol, and all associated code (available at <https://github.com/mleyvaz/woodall-response-experiments>) are offered as tools for any researcher to test these claims on their own data.

References

- Aslam, M., AlAita, A., & Smarandache, F. (2025). A critical evaluation of the criticisms against neutrosophic statistical methods. *Neutrosophic Systems with Applications*, 25(6), 576–589. <https://doi.org/10.63689/2993-7159.1283>
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7, 1–30.
- Haq, A., & Woodall, W. H. (2025). A critique of neutrosophic statistical analysis illustrated with interval data from designed experiments. *Journal of Quality Technology*. [Submitted].
- Moore, R. E. (1966). *Interval analysis*. Prentice-Hall.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.
- Smarandache, F. (1998). *Neutrosophy: Neutrosophic probability, set, and logic*. American Research Press.

- Smarandache, F. (2014). Introduction to neutrosophic statistics. Sitech & Education Publishing.
<http://fs.unm.edu/NeutrosophicStatistics.pdf>
- Smarandache, F. (2021). Plithogenic probability & statistics are generalizations of multivariate probability & statistics. *Neutrosophic Sets and Systems*, 43, 280–289.
- Smarandache, F. (2022). Neutrosophic statistics is an extension of interval statistics, while plithogenic statistics is the most general form of statistics (second version). *International Journal of Neutrosophic Science*, 19(1), 148–165. <https://doi.org/10.54216/IJNS.190111>
- Steiner, S. H., & Woodall, W. H. (2025). A simulation-based approach for process monitoring with interval-valued data. *Quality Engineering*. Advance online publication.
- Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer.
- Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman & Hall.
- Woodall, W. H., King, C., Driscoll, A. R., & Montgomery, D. C. (2025). A critical assessment of neutrosophic statistical methods. *Quality Engineering*. <https://doi.org/10.1080/08982112.2025.2482198>